



별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto is a true copy from the records of the Korean Intellectual Property Office.

출원 번호 : 10-2003-0050552
Application Number

출원 년 월 일 : 2003년 07월 23일
Date of Application JUL 23, 2003

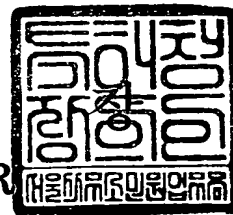
출원인 : 주식회사 팬택
Applicant(s) PANTECH CO., LTD.



2003 년 12 월 04 일

특 허 청

COMMISSIONER





【서지사항】

【서류명】	특허출원서		
【권리구분】	특허		
【수신처】	특허청장		
【제출일자】	2003.07.23		
【발명의 명칭】	HMM 확률 보정 방법		
【발명의 영문명칭】	METHOD FOR MODIFICATING HMM		
【출원인】			
【명칭】	주식회사 팬택		
【출원인코드】	1-1998-004053-1		
【대리인】			
【명칭】	특허법인 신성		
【대리인코드】	9-2000-100004-8		
【지정된변리사】	변리사 신윤정, 변리사 원석희, 변리사 박해천		
【포괄위임등록번호】	2002-089790-8		
【발명자】			
【성명의 국문표기】	권태희		
【성명의 영문표기】	KWON, Tae Hee		
【주민등록번호】	751030-1496210		
【우편번호】	570-080		
【주소】	전라북도 익산시 송학동 현대아파트 105-303		
【국적】	KR		
【공지예외적용대상증명서류의 내용】			
【공개형태】	간행물 발표		
【공개일자】	2003.02.28		
【심사청구】	청구		
【취지】	특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사를 청구합니다. 대리인 특허법인 신성 (인)		
【수수료】			
【기본출원료】	20	면	29,000 원
【가산출원료】	4	면	4,000 원



1020030050552

출력 일자: 2003/12/13

【우선권주장료】	0	건	0	원
【심사청구료】	7	항	333,000	원
【합계】	366,000			원
【첨부서류】	1. 요약서·명세서(도면)_1통 2. 공지에외적용대상(신규성상실의예 외, 출원시의특례)규정을 적용받 기 위한 증명서류_1통			

【요약서】**【요약】**

본 발명의 HMM 확률 보정 방법은, 훈련 데이터에 대한 인식 오류를 감소시키기 위하여 최적 상태열을 구하는 과정 및 경도 계산 과정을 생략하고, 음성 인식 단위 집합을 구성하는 각각의 HMM에 대응하는 가중치를 도입함으로써, 파라미터를 감소시키는 동시에 경도 계산의 미수행에 의한 계산량의 감소에 따른 음성 인식 성능 향상을 도모할 수 있는 HMM 확률 보정 방법을 제공하는데 그 목적이 있다.

상기 목적을 달성하기 위하여 본 발명은, 분별함수를 통하여 입력되는 음성의 패턴을 인식하는 단계; 입력 음성에 대하여 클래스를 결정하는 단계; 상기 클래스의 결정에 따른 음성 인식 성능을 나타내는 평균손실함수를 획득하는 단계; 상기 평균손실함수에 의해 모델 파라미터를 도출하는 단계; 및 상기 모델 파라미터 식에 따라 HMM 가중치 훈련을 수행하는 단계를 포함하고, 상기 분별함수는 각 클래스에 대한 HMM 가중치를 상기 분별함수 내 상태 천이 성분 및 확률 밀도 성분에 적용하는 것을 특징으로 한다.

【대표도】

도 1

【색인어】

HMM 가중치, 클래스, 음성 인식, 분별 함수

【명세서】**【발명의 명칭】**

HMM 확률 보정 방법{METHOD FOR MODIFICATING HMM}

【도면의 간단한 설명】

도 1은 본 발명의 일 실시예에 의한 HMM 확률 보정 방법을 나타낸 동작흐름도,

도 2는 본 발명의 일 실시예에 의한 HMM 확률 보정 방법에 적용되는 모델 파라미터 식의 HMM 가중치 추정 방법을 나타낸 동작흐름도,

도 3은 HMM 가중치 훈련을 나타낸 그래프.

【발명의 상세한 설명】**【발명의 목적】****【발명이 속하는 기술분야 및 그 분야의 종래기술】**

- <4> 본 발명은 HMM 확률 보정 방법에 관한 것으로, 특히, HMM 확률 보정을 이용하여 음성 인식 성능을 향상시키는 HMM 확률 보정 방법에 관한 것이다.
- <5> 일반적으로, 은닉 마코프 모델링(Hidden Markov Modeling)은 음성 인식에서 음성의 특징을 표현하는데 매우 널리 사용되고 있다. 은닉 마코프 모델링은 통계적인 기반 위에서 음성의 특징을 모델링하는 일을 HMM 상태의 확률 분포 추정 방법으로 귀결시킨다. 이러한 상태 확률 분포 추정 방법으로써 가장 널리 사용되는 방법은 ML(Maximum Likelihood) 추정 방법이다.

- <6> 그러나, ML 추정 방법은 훈련 데이터의 확률 분포에 대한 정확한 정보를 알아내는 일이 매우 어려우며 유사 음성들 사이의 구분되는 정보를 모델링하기가 거의 불가능하다. 패턴 인식 문제에서 위와 같은 확률 분포 추정의 단점을 보완하기 위한 방법으로 분별 훈련 방법이 많이 연구되어 왔다.
- <7> 특히, 음성 인식기의 성능은 인식기의 평균 인식 에러율로 정의되며 최적 인식기는 최소의 인식 에러율을 나타내는 인식기이다. 이러한 관점에서 주로 연구되어 온 분별 훈련 기법이 GPD(Generalized Probabilistic Descent) 알고리즘에 기반한 MCE(Minimum Classification Error) 훈련 방법이다.
- <8> MCE 훈련 기법의 목적은 데이터의 확률 분포를 추정하여 모델을 얻는 것이 아니고 최고의 인식 결과를 위해서 HMM의 관측 데이터를 구분하는 것이다. 이하 MCE 알고리즘을 간략히 설명하고 에러 확률의 합리적인 추정치를 나타내는 최적화 범주에 대하여 논의한다. 또한, MCE 훈련 방법은 궁극적으로 인식 오류의 최소화 관점에서 접근하며 특징 추출, 음향 모델링 기법, 음향 모델의 정밀도가 고정된 상태에서 믹스처(mixture) 가중치, 평균, 표준편차 등의 HMM 기본 파라미터들을 조절해 줌으로써 음성 인식의 성능 향상을 달성할 수 있음이 이미 연구되어 왔다. MCE 훈련 방법의 확장된 방법으로써 HMM 상태 확률에 내재해 있는 음성들 사이의 구분되는 정보를 이용하여 인식기를 최적화하기 위해서 상태 가중치를 도입하는 방법이 연구되었다. MCE 훈련 기법은 주로 ML 훈련 기법과 병행하여 행해지고 있으며 ML 훈련 방법에 의해 추정된 HMM보다 우수한 성능을 보인다.
- <9> 기본적인 HMM 기반의 음성 인식기에서, 패턴 인식을 위해 클래스 i 에 대한 분별 함수는 다음 식과 같이 정의된다.

<10>

$$g_i(\mathbf{X}; \Lambda) = \log \{g_i(\mathbf{X}, \bar{\mathbf{q}}; \Lambda)\}$$

$$= \sum_{t=1}^T \left[\log a_{\bar{q}_{t-1}\bar{q}_t}^{(i)} + \log b_{\bar{q}_t}^{(i)}(\mathbf{x}_t) \right] + \log \pi_{\bar{q}_0}^{(i)}$$

【수학식 1】

<11>

여기에서, $\bar{\mathbf{q}} = (\bar{q}_0, \bar{q}_1, \Lambda, \bar{q}_T)$ 는, 클래스 i 에 대해 조인트 상태열-관측열 확률 함수 $g_i(\mathbf{X}, \bar{\mathbf{q}}; \Lambda)$ 를 최대화하는 최적의 상태열이고, a_{ij} 는 상태 i 에서 상태 j 로 천이하는 상태 천이 확률을 의미한다.

<12>

$b_j(\mathbf{x}_t)$ 는 상태 j 에서 관측벡터 \mathbf{x}_t 를 관측하는 확률 밀도 함수를 나타낸다. 연속적인 다변수 믹스처(mixture) 가우시안 HMM에서는, 상태 출력 분포는 다음과 같다.

<13>

$$b_j(\mathbf{x}_t) = \sum_{m=1}^M c_{jm} N(\mathbf{x}_t; \mu_{jm}, \Sigma_{jm})$$

【수학식 2】

<14>

여기에서, $N(\cdot)$ 은 다변수 가우시안 밀도 함수를 표기하며 μ_{jm} 는 상태 j , 믹스처(mixture) m 에서의 평균 벡터이고, Σ_{jm} 는 상태 j , 믹스처(mixture) m 에서의 상관행렬이다.

<15>

입력 음성에 대하여, 클래스 결정 규칙이 사용되며 입력 음성 X 에 대한 클래스 $C(X)$ 는 다음과 같은 규칙에 의해 결정된다.

<16>

$$C(\mathbf{X}) = C_i \quad \text{if } i = \arg \max_j g_j(\mathbf{X}; \Lambda)$$

【수학식 3】

<17>

여기에서, C_i 는 분별함수에 의해 입력 음성 또는 관측 벡터열에 대해 결정된 클래스를 의미한다.

<18> 먼저, 연산적인 결정 규칙인 수학식 3을 함수 형태로 표현하는 것이 필요하다. 패턴 인식기의 파라미터 세트 Λ 의 연속적인 함수로서 결정 규칙을 함축하는 클래스 오인식 척도는 다음과 같이 정의된다.

<19>

$$d_i(\mathbf{X}; \Lambda) = -g_i(\mathbf{X}; \Lambda) + \log \left[\frac{1}{N} \sum_{j=1, j \neq i}^N \exp \left[g_j(\mathbf{X}; \Lambda) \eta \right] \right]^{\frac{1}{\eta}}$$

【수학식 4】

<20> 여기에서, η 는 양의 상수이고 N 은 N -best 오인식 클래스들의 개수이다. 클래스 i 에 해당하는 음성 X 에 대해, $d_i(\mathbf{X}) \geq 0$ 는 오인식을 의미하며 $d_i(\mathbf{X}) < 0$ 는 정확한 인식을 의미한다.

<21> 완전한 손실함수는 부드러운 이진 손실함수의 형태로서 오인식 척도에 관하여 정의된다.

<22>

$$l_i(\mathbf{X}; \Lambda) = l(d(\mathbf{X}))$$

【수학식 5】

<23> 부드러운 이진 손실함수는 임의의 연속적인 이진 함수로 정의될 수 있으나 보통 다음과 같은 S형(sigmoid) 함수가 사용된다.

<24>

$$l(d) = \frac{1}{1 + \exp[-rd + \theta]}$$

【수학식 6】

<25> 여기에서, θ 는 영 또는 영보다 다소 작은 값으로 설정되고 r 은 상수값이다.

<26> 마지막으로, 미지의 음성에 대하여 인식기 성능은 평균 손실함수로서 평가된다.

<27>

$$l(\mathbf{X}; \Lambda) = \sum_{i=1}^M l_i(\mathbf{X}; \Lambda) l(\mathbf{X} \in C_i)$$

【수학식 7】

<28> 여기에서, $l(\cdot)$ 는 표시자(indicator) 함수이다.

<29> 최적의 모델 파라미터는 평균 손실을 최소로 하는 모델 파라미터이며 평균 손실을 최소화하기 위해서 GPD 알고리즘이 주로 사용된다. GPD 알고리즘은 다음과 같이 주어진다.

<30>

$$\Lambda_{n+1} = \Lambda_n - \varepsilon_n \mathbf{U}_n \nabla l(\mathbf{X}; \Lambda) |_{\Lambda = \Lambda_n}$$

【수학식 8】

<31> 여기에서, \mathbf{U} 는 양으로 정의된 행렬, ε_n 는 학습 비율 또는 조절의 스텝 크기(step size)이고, Λ_n 는 시각 n 에서 모델 파라미터 세트이다.

<32> GPD 알고리즘은 제한 조건이 없는 최적화 기술이다. 그러나, 확률 모델로서의 HMM 구조를 유지하기 위해서는 어떠한 제약 조건이 주어져야만 한다. 복잡한 제약조건을 갖는 GPD 알고리즘을 사용하는 대신에 GPD 알고리즘을 변환된 HMM 파라미터에 적용하였다. 파라미터 변환 과정은 변환된 공간에서 어떠한 제약조건이 없으며 원시 공간으로의 변환시 HMM 제약조건이 만족되어야 한다. 다음과 같은 HMM 파라미터에 대한 제약조건이 원시 공간에서 유지되어야 한다.

<33>

$$\sum_j a_{ij} = 1 \text{ and } a_{ij} \geq 0, \sum_k c_{jk} = 1 \text{ and } c_{jk} \geq 0, \sigma_{jkl} \geq 0$$

【수학식 9】

<34> 상술한 식과 같은 원시 공간에서의 파라미터 제약 조건을 만족시키기 위해서 다음 식과 같은 파라미터 변환 과정이 파라미터 훈련 전후에 사용된다.

<35>

$$\begin{aligned} \dot{a}_{ij} &\rightarrow \tilde{a}_{ij} \quad \text{where } a_{ij} = e^{\tilde{a}_{ij}} / \left(\sum_k e^{\tilde{a}_{ik}} \right) \\ c_{jk} &\rightarrow \tilde{c}_{jk} \quad \text{where } c_{jk} = e^{\tilde{c}_{jk}} / \left(\sum_k e^{\tilde{c}_{jk}} \right) \\ \mu_{jkl} &\rightarrow \tilde{\mu}_{jkl} = \mu_{jkl} / \sigma_{jkl} \\ \sigma_{jkl} &\rightarrow \tilde{\sigma}_{jkl} = \log \sigma_{jkl} \end{aligned}$$

【수학식 10】

<36>

그러나, segmental GPD 알고리즘에 기반한 MCE 훈련은 훈련 클래스에 의한 최적 상태열을 구하는 과정과 HMM을 구성하는 파라미터에 대한 경도 계산이 요구되며, 이러한 과정은 그 계산량이 많이 소요되어 연속어를 이용한 훈련에 적용하기에 어려운 문제점이 있다.

【발명이 이루고자 하는 기술적 과제】

<37>

상기 문제점을 해결하기 위하여 안출된 본 발명은, 훈련 데이터에 대한 인식 오류를 감소시키기 위하여 최적 상태열을 구하는 과정 및 경도 계산 과정을 생략하고, 음성 인식 단위 집합을 구성하는 각각의 HMM에 대응하는 가중치를 도입함으로써, 파라미터를 감소시키는 동시에 경도 계산의 미수행에 의한 계산량의 감소에 따른 음성 인식 성능 향상을 도모할 수 있는 HMM 확률 보정 방법을 제공하는데 그 목적이 있다.

【발명의 구성 및 작용】

<38>

상기 목적을 달성하기 위하여 본 발명의 HMM 확률 보정 방법은, 분별함수를 통하여 입력되는 음성의 패턴을 인식하는 단계; 입력 음성에 대하여 클래스를 결정하는 단계; 상기 클래스의 결정에 따른 음성 인식 성능을 나타내는 평균손실함수를 획득하는 단계; 상기 평균손실함수

에 의해 모델 파라미터식을 도출하는 단계; 및 상기 모델 파라미터 식에 따라 HMM 가중치 훈련을 수행하는 단계를 포함하고, 상기 분별함수는 각 클래스에 대한 HMM 가중치를 상기 분별함수 내 상태 천이 성분 및 확률 밀도 성분에 적용하는 것을 특징으로 한다.

- <39> 이하, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 본 발명의 기술적 사상을 용이하게 실시할 수 있을 정도로 상세히 설명하기 위하여 본 발명의 가장 바람직한 실시예들을 첨부된 도면을 참조하여 설명하기로 한다.
- <40> 도 1은 본 발명의 일 실시예에 의한 HMM 확률 보정 방법을 나타낸 동작흐름도로서, 이러한 본 발명의 HMM 확률 보정 방법에 관하여 설명하면 다음과 같다.
- <41> 먼저, 분별함수를 통하여 입력되는 음성의 패턴을 인식한다(S110). 여기서, 유사한 음성을 구분해 주기 위해서 HMM 출력 확률에 내재해 있는 구별되는 정보를 이용하기 위해서 각각의 HMM에 가중치를 주어 오인식 척도가 영이 되는 방향으로 반복하여 훈련함으로써 음성 인식 오류의 감소를 기대할 수 있다. 이를 위해서 HMM의 출력 스코어는 비터비(Viterbi) 탐색 과정에 의한 HMM 출력 확률값과 HMM의 가중치의 곱으로 표현한다. 수식적인 설명을 위하여 단어 모델을 이용한 고립 단어 인식의 경우를 가정하여 기본적인 음성 인식 단위로서 M개의 HMM을 설정하고 각각의 인식 단위는 J개의 상태로 구성된다고 가정하자. HMM 기반의 패턴 인식은 클래스 i에 대한 분별 함수를 이용한 클래스 결정 규칙에 의해 패턴 인식이 이루어지게 된다. 이와 같은 패턴 인식 과정에 있어서 클래스 i에 대한 분별 함수는 수식식 1에 의해 정의된다. 이와 유사하게 확장된 HMM 기반의 패턴 인식 과정에 있어서 클래스 i에 대한 분별 함수는 다음 식에 의해 정의된다.

<42>

$$\begin{aligned}
 g_i(\mathbf{X}; \Lambda) &= w_i \left[\sum_{t=1}^T \left\{ \log a_{\bar{q}_{t-1}\bar{q}_t}^{(i)} + \log b_{\bar{q}_t}^{(i)}(x_t) \right\} + \log \pi_{\bar{q}_0}^{(i)} \right] \\
 &= \sum_{t=1}^T \left\{ w_i \cdot \log a_{\bar{q}_{t-1}\bar{q}_t}^{(i)} + w_i \cdot \log b_{\bar{q}_t}^{(i)}(x_t) \right\} + w_i \cdot \log \pi_{\bar{q}_0}^{(i)}
 \end{aligned}$$

【수학식 11】

<43>

여기에서, w_i 는 클래스 i 에 대한 HMM 가중치이다.

<44>

인식에 사용되는 HMM 집합에 대한 HMM 가중치의 합은 다음 식과 같이 HMM의 총 개수로 제한되어야 한다.

<45>

$$\sum_{i=1}^M w_i = M, \quad 0 < w_i < M$$

【수학식 12】

<46>

HMM 가중치가 1로 설정되었을 때 N-best 스트링 모델을 기반으로 한 인식 알고리즘이 동일한 결과를 도출하도록 HMM 가중치에 대한 제한 조건을 주는 이유는 기존의 파라미터 추정 방법과 비터비 탐색 알고리즘에 의한 확률값의 큰 변화 없이 용이하게 인식 과정을 수행하기 위함이다.

<47>

또한, 보통 훈련 데이터에 대한 인식 오류의 개수를 감소시키기 위한 방법으로 N-best 스트링에 기반한 세그멘탈(Segmental) GPD 알고리즘이 사용된다. 그러나, 경도 계산에 의한 계산량 증가를 피하기 위하여 HMM 가중치에 대한 손실 함수의 미분값을 계산하는 대신에 우리는 훈련 데이터에 대한 오인식 척도를 감소시키기 위해서 요구되는 HMM 가중치의 변화량을 도입하였다. 이러한 HMM 가중치의 변화량은 간단한 나눗셈 연산을 통하여 매우 간단하게 구할 수 있다.

<48> 훈련 클래스 모델과 경쟁하는 N개의 클래스 모델과의 확률적 거리 척도는 이후로부터 기본 식과 훈련 음성의 스트링의 확률값에 가중치를 주어 더해주는 확장된 식으로 두 가지가 사용될 수 있으며 다음 식과 같다.

<49>

$$d_i(\mathbf{X}; \Lambda) = -g_i(\mathbf{X}; \Lambda) + \log \left[\frac{1}{N} \sum_{j=1, j \neq i}^N \exp[g_j(\mathbf{X}; \Lambda)\eta] \right]^{\frac{1}{\eta}}$$

$$\tilde{d}_i(\mathbf{X}; \Lambda) = d_i(\mathbf{X}; \Lambda) - k \cdot g_i(\mathbf{X}; \Lambda)$$

$$= -(1+k) \cdot g_i(\mathbf{X}; \Lambda) + \log \left[\frac{1}{N} \sum_{j=1, j \neq i}^N \exp[g_j(\mathbf{X}; \Lambda)\eta] \right]^{\frac{1}{\eta}}$$

【수학식 13】

<50> 클래스 i에 대한 HMM 가중치 조절을 위해서는 클래스 i에 대한 HMM 가중치가 변화되어야 하는 비율을 설정해야 하며 이러한 비율을 델타(delta) 계수라고 정의하고 Δw_i 로 표기한다. 클래스 i에 대한 HMM 가중치 조절을 위한 델타 계수는 오인식 척도와 훈련 음성의 클래스 i에 대한 분별 함수의 값을 이용하여 다음 식과 같이 정의한다.

<51>

$$\Delta w_i = \frac{d_i(\mathbf{X}; \Lambda)}{-g_i(\mathbf{X}; \Lambda)}$$

【수학식 14】

<52> 그 후, 입력 음성에 대하여 클래스 결정규칙을 적용한다(S120). 즉, 입력 음성 또는 관측벡터열에 대해 클래스를 결정한다.

<53> 그 후, 상기 클래스의 오인식에 따라 음성 인식 성능을 나타내는 평균손실함수를 획득한다(S130).



<54> 그 후, 상기 평균손실함수에 의해 모델 파라미터식을 도출한다(S140).

<55> 초기값 1로 설정된 클래스 i에 대한 HMM 가중치의 훈련은 클래스 오인식 척도를 감소시키기 위해 요구되는 클래스 i에 대한 HMM 가중치에 대한 델타 계수 Δw_i 를 이용하여 다음과 같은 모델 파라미터식에 따라 반복적으로 훈련한다.

<56>
$$\tilde{w}_i(n+1) = w_i(n) - \varepsilon_n \cdot w_i(n) \cdot \Delta w_i$$

 【수학식 15】

<57> 끝으로, 상기 모델 파라미터 식에 따라 HMM 가중치 훈련을 수행하고(S150), 상기 HMM 가중치 훈련 전후에 파라미터 변환을 수행하는데, 상기 파라미터 변환은 다음 식에 의해 주어진다.

<58> HMM 집합에 대한 HMM 가중치의 합이 HMM 집합 내의 HMM 총 개수가 되는 HMM 가중치에 대한 제약 조건을 만족시키기 위해서 다음 식과 같이 파라미터 변환 과정을 HMM 가중치에 적용한다.

<59>
$$w_i \rightarrow \tilde{w}_i \quad \text{where } w_i = e^{\tilde{w}_i} / \left(\sum_k e^{\tilde{w}_k} \right)$$

 【수학식 16】

<60> 여기에서, \tilde{w}_i 는 원시 공간에서의 클래스 i에 대한 HMM 가중치 w_i 에 대응하는 변환된 공간에서의 HMM 가중치이다.

<61> 또한, 연속 음성 인식을 위한 인식 알고리즘은 비터비 탐색 과정에서 각각의 HMM 가중치를 고려하여 계산함으로써 다음과 같이 구현될 수 있다.

<62>

$$V[0][j] = 0, j = \pi_0$$

$$V[0][j] = -\infty, j \neq \pi_0$$

$$V[t][j] = \max_h \left[V[t-1][h] + w(h) \cdot \{\log a_{hj}\} + w(j) \cdot \log b_j(x_t) \right]$$

【수학식 17】

$$w(j) = w_k \text{ if } j \in H_k, k = 1, 2, \Lambda, M$$

<63>

여기에서, $V[t][j]$ 는 시간 t 에서 상태 j 까지의 누적된 스코어이고, π_0 는 초기 상태를 의미하며 H_k 는 k 번째 HMM을 의미하고 $\log b_j(x_t)$ 는 상태 j 에서 관측 벡터 x_t 를 관측할 log 확률값을 의미하며 w_k 는 k 번째 HMM에 대한 HMM 가중치를 의미한다.

<64>

도 2는 본 발명의 일 실시예에 의한 HMM 확률 보정 방법에 적용되는 모델 파라미터 식의 HMM 가중치 추정 방법을 나타낸 동작흐름도로서, 이에 관하여 설명하면 다음과 같다.

<65>

훈련 음성에 대한 클래스 i 가 K 개의 HMM으로 구성되어 있다고 가정하자. 먼저, 인식 대상인 음성을 입력받아(S210), 연속 음성 인식을 위한 비터비 디코딩을 수행하고(S220), N-best 탐색 과정에 의해 구한 오인식 척도가 양인지 여부를 판단한다(S230).

<66>

만약, 오인식 척도가 양인 경우 오인식 척도가 감소하도록 훈련 클래스에 대한 스코어를 조절해 주기 위해서 델타(delta) 계수를 구하고(S240), 이를 훈련 클래스를 구성하는 개개의 HMM 가중치에 반영하게 된다(S250). 이후에는, HMM 가중치가 반영된 비터비 디코딩 과정이 수행된다. 따라서, 하기 식과 같은 과정에 의해 훈련 클래스를 구성하는 HMM 가중치를 훈련한다.

<67>

【수학식 18】

$$\tilde{w}_k^{(i)}(n+1) = w_k^{(i)}(n) - \varepsilon_n \cdot w_k^{(i)}(n) \cdot \Delta w_i, k = 1, 2, \Lambda, K$$

<68> 여기에서, $w_k^{(i)}$ 는 클래스 i를 구성하는 k번째 HMM에 대한 가중치이고, Δw_i 는 클래스 i에 대한 델타(delta) 계수이고, ε_n 은 n번째 반복 훈련 과정에서의 학습 비율이다.

<69> 이와 같이 훈련된 HMM 가중치는 하기 식에 의해 주어지는 HMM 가중치에 대한 제약 조건을 만족시키도록 파라미터 변환 과정에 의해서 조정한다.

<70>

$$w_k \rightarrow \tilde{w}_k \quad \text{where } w_k = e^{\tilde{w}_k} / \left(\sum_{x=1}^M e^{\tilde{w}_x} \right)$$

【수학식 19】

<71> 한편, 오인식 척도가 양이 아닌 경우에는, 새로운 음성을 받아들이기 위해 인식 대상인 음성을 입력받는 단계(S210)로 돌아가게 된다.

<72> HMM 확률 보정을 통한 음성 인식의 성능 향상을 평가하기 위하여 단독 숫자음을 이용한 실험을 실시하였다. 단독 숫자음 DB는 조용한 사무실 환경에서 녹음되었으며 500명(남성 250명, 여성 250명)의 음성으로 구성되었고, 400명(남성 200명, 여성 200명)의 음성이 훈련에 사용되었고 100명(남성 50명, 여성 50명)의 음성이 인식에 사용되었다. 본 실험에서는 11.025KHz DSP 보드 상에서의 음성 인식기 구현을 위해서 16KHz로 샘플링된 음성을 11.025KHz로 다운 샘플링하였다. 특징 벡터로는 log 에너지를 포함하여 13차 cepstral 계수, 13차 1차 미분계수, 13차 2차 미분 계수를 이용하여 총 3스트림, 39차 특징 벡터를 구성하였다. 모든 상태 출력 확률 분포는 8 믹스처(mixture) 다변수 가우시안 분포를 사용하였다. 베이스 라인 실험으로써 ML 훈련 방법에 의해서 3개의 상태, 3개의 스트림, 8개의 믹스처(mixture)로 구성된 모델을 생성하였다. HMM 가중치를 훈련하는 과정에서 오인식 척도를 계산하기 위해 3개의 가장 경쟁적인 스트링을 사용하였다(N=3).

<73> 본 실험에서는 먼저 4가지의 손실 함수(오인식 척도에 기인한 이진 손실 함수, 확장된 오인식 척도에 기인한 이진 손실 함수, 오인식 척도, 확장된 오인식 척도)를 사용하여 HMM 기본 파라미터를 MCE 훈련하여 인식 실험을 실시하였다. 하기 표 1은 훈련 데이터 인식률이 최대가 되었을 경우 훈련 및 인식 데이터에 대한 단어 인식률을 나타낸 것이다. 훈련을 반복함에 따라 훈련 데이터 인식률은 점차적으로 증가하나 인식 데이터에 대한 인식률은 거의 변화가 없는데, 이는 셉트럴(cepstral) 계수에 대한 상태 확률 분포가 훈련 데이터에 과적응 되어가는 과정임을 의미한다.

<74> 【표 1】

Data	MLE	SIG	SIG+WL	LIN	LIN+WL
Training	98.40	98.58	98.58	98.58	98.71
Testing	98.00	98.00	98.00	98.00	98.00

SIG : sigmoid loss function

LIN : linear loss function

WL : weighted likelihood of correct class

<75> 도 3은 확장된 오인식 척도(LIN+WL)를 사용하여 HMM 가중치를 훈련하는 HMM 확률 보정 기법에 따라 훈련 데이터를 이용하여 15번 반복하여 훈련함에 따른 훈련 데이터와 인식 데이터의 변화 과정을 보여준다. MCE 훈련 방법의 원리와 동일하게 반복 회수가 증가함에 따라 훈련 데이터와 인식 데이터에 대한 인식률이 일관성 있게 상승됨을 보여준다. 도 3의 훈련 데이터 인식률이 인식 데이터에 대한 인식률보다 높게 나타나는 경우는 훈련 데이터에 부정확한 발음

이 포함된 데에 기인하며 이러한 음성은 HMM 상태 가중치 훈련에 사용하지 않았다. HMM 확률 보정 기법은 상태 확률 보정 기법에 비하여 훈련 데이터에 대한 과적응에 덜 민감한 특성을 보이며 오인식 척도 자체를 델타 계수를 이용하여 HMM 가중치에 직접 반영함으로써 가중치 훈련을 보다 효과적으로 할 수 있는 장점이 있다.

<76> HMM 기본 파라미터의 MCE 훈련, 상태 확률 보정 기법, HMM 확률 보정 기법의 세가지 실험을 실시하여 인식 데이터 인식률이 최고일 경우 훈련 및 인식 데이터 인식률을 표 2에 나타내었다. 하기 표 2에서 볼 수 있듯이, HMM 확률 기법은 최적 상태열 및 경도 계산을 수행하지 않고도 가장 높은 인식 성능 향상을 보여준다.

<77> 【표 2】

Data	ML	MCE	State Weighting	HMM Weighting
Training	98.40	98.71	98.56	98.81
Testing	98.00	98.00	98.50	98.92

<78> 이상에서 설명한 본 발명은, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 있어 본 발명의 기술적 사상을 벗어나지 않는 범위 내에서 여러 가지로 치환, 변형 및 변경이 가능하므로 전술한 실시예 및 첨부된 도면에 한정되는 것이 아니다.

【발명의 효과】

<79> HMM 기반의 음성 인식 시스템에서 훈련 음성에 대한 훈련 음소열의 확률값에 가중치를 주어 얻어진 확장된 선형 손실 함수를 사용하여 HMM 가중치를 훈련함으로써 훈련 및 음성 데이터에 대한 인식 성능이 일관성 있게 증가함을 보여준다. 또한, 제안된 방법인 HMM 가중치를 델타(delta) 계수값을 이용하여 반복적으로 훈련함으로써 기본 파라미터의 MCE 훈련 방법 및 상태 확률 보정 기법에 비해 더욱 우수한 음성 인식 성능 향상을 얻을 수 있는 장점이 있다.



【특허청구범위】

【청구항 1】

분별함수를 통하여 입력되는 음성의 패턴을 인식하는 단계;

입력 음성에 대하여 클래스를 결정하는 단계;

상기 클래스의 결정에 따른 음성 인식 성능을 나타내는 평균손실함수를 획득하는 단계;

상기 평균손실함수에 의해 모델 파라미터식을 도출하는 단계; 및

상기 모델 파라미터 식에 따라 HMM 가중치 훈련을 수행하는 단계

를 포함하고,

상기 분별함수는 각 클래스에 대한 HMM 가중치를 상기 분별함수 내 상태 천이 성분 및 확률 밀도 성분에 적용하는

것을 특징으로 하는 HMM 확률 보정 방법.

【청구항 2】

제1항에 있어서,

상기 클래스의 결정에 따른 음성 인식 성능을 나타내는 평균손실함수를 획득하는 단계는,

훈련 음성의 스트링의 확률값에 가중치를 주어 오인식 척도에 더해서 획득되는 확장된 오인식 척도를 사용하는

것을 특징으로 하는 HMM 확률 보정 방법.

【청구항 3】

제1항에 있어서,
상기 HMM 가중치의 훈련의 초기값은 1로 설정되는
것을 특징으로 하는 HMM 확률 보정 방법.

【청구항 4】

제1항에 있어서,
상기 HMM 가중치의 합은 HMM의 총 개수로 제한되는
것을 특징으로 하는 HMM 확률 보정 방법.

【청구항 5】

제1항에 있어서,
상기 분별함수를 통하여 입력되는 음성의 패턴을 인식하는 단계는,
비터비 탐색 과정에서 각각의 상기 HMM 가중치를 고려하여 계산하는 단계
를 포함하는 것을 특징으로 하는 HMM 확률 보정 방법.

【청구항 6】

제1항에 있어서,

상기 HMM 가중치는, 훈련 클래스 모델 및 훈련 클래스 모델과 경쟁하는 클래스 모델과의 확률적 거리 척도인 오인식 척도를 훈련 클래스의 분별함수로 나눈 델타 계수를 이용하여 조정되는

것을 특징으로 하는 HMM 확률 보정 방법.

【청구항 7】

제1항에 있어서,

인식 대상인 음성을 입력받는 단계;

HMM 가중치에 따른 연속 음성 인식을 위한 비터비 디코딩을 수행하는 단계;

오인식 척도가 양인지 여부를 판단하고, 오인식 척도가 양이 아닌 경우에는, 새로운 음성을 받아들이기 위해 인식 대상인 음성을 입력받는 단계로 돌아가는 단계;

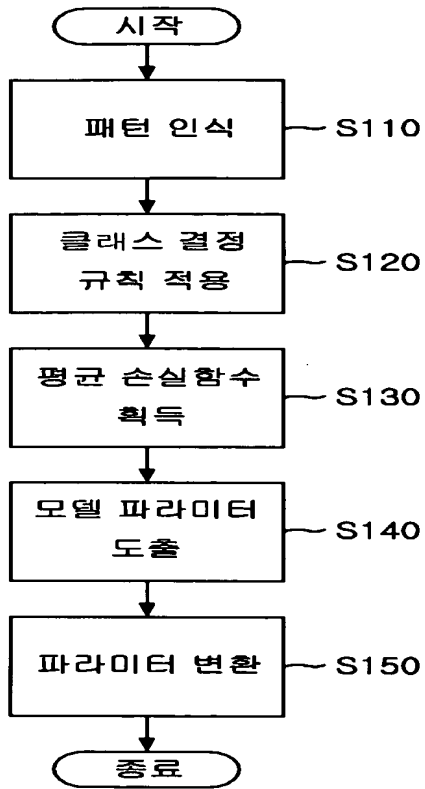
오인식 척도가 양인 경우 상기 오인식 척도가 감소하도록 조정하는 단계; 및

상기 오인식 척도의 감소를 상기 HMM 가중치에 반영하는 단계

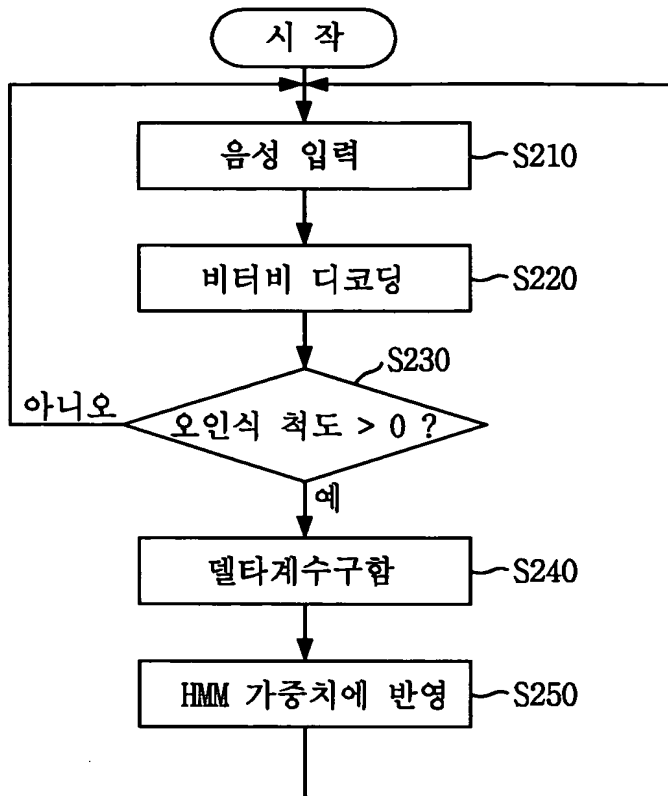
를 포함하는 것을 특징으로 하는 HMM 확률 보정 방법.

【도면】

【도 1】



【도 2】



【도 3】

